ADL hw2 report

R13944037

Q1. Model

Architecture:

The google/mt5-small model is a variant of the Multilingual Text-to-Text Transfer Transformer (mT5), designed to handle a wide array of natural language processing tasks, including text summarization. As a text-to-text transformer model, it treats all tasks as text generation problems, allowing it to convert input text into output text across various applications such as translation and summarization. Like its predecessor T5, mT5-small utilizes a transformer architecture with an encoder-decoder structure, enabling effective processing and generation of text by encoding input and decoding it into the desired output format.

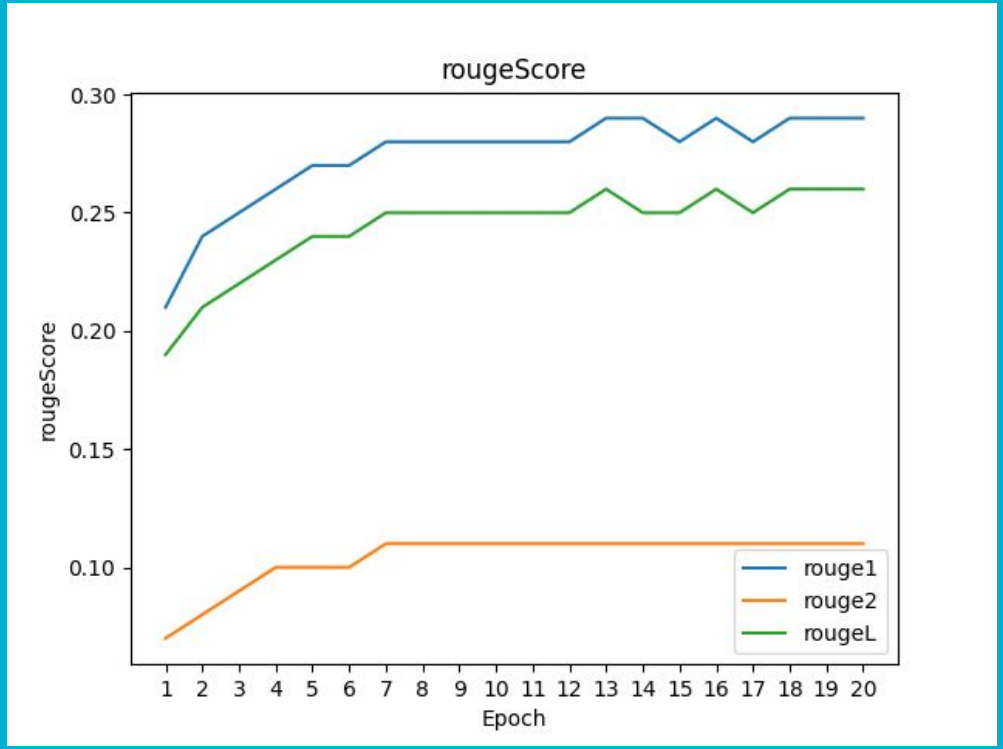
Preprocess:

For every input, I prepend 'summarize:' to specify the task. I used the mT5 pre-trained tokenizer.

Q2. Training

```
--per_device_train_batch_size 2 \
--gradient_accumulation_steps 8 \
--num_train_epochs 20 \
--model_name_or_path google/mt5-small \
--learning_rate  3e-4 \
--max_target_length 1024 \
--max_source_length 2048 \
--source_prefix "summarize: " \
```

Q2. Training


rougeScore
0.30
0.25
0.20
0.15
0.10
rougeScore
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Epoch
rouge1
rouge2
rougeL

Q3: Generation Strategies

- Greedy Search: Selects the most probable word at each step, often leading to fast but suboptimal outputs.
- Beam Search: Maintains a fixed number of the best candidate sequences at each step, improving overall quality but requiring more computation.
- Top-k Sampling: Randomly selects from the top k most probable words, introducing diversity while controlling for coherence.
- Top-p Sampling: Chooses from a dynamic set of words whose cumulative probability exceeds a threshold p, balancing relevance and randomness.
- Temperature: Adjusts the randomness of predictions, with lower values making outputs more deterministic and higher values increasing creativity.

Q3: Generation Strategies

	Greedy	num_beams = 5
rouge-1	0.2646	0.2775
rouge-2	0.1009	0.1134
rouge-l	0.2357	0.2475